DOCUMENT RESUME

ED 166 974 FL 010 894

AUTHOR Esau, Helmut; Yost, Carlson
TITLE Measuring Writing Ability with the Cloze Test is not Closed.
PUB DATE 78
NOTE 11p.; Paper presented at the meeting of the Linguistic Society of America (Urbana, Illinois, Summer 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Cloze Procedure; College Freshmen; *Composition Skills (Literary); English; *English Instruction; Essay Tests; Higher Education; Language Ability; Language Skills; Testing; *Tests; *Verbal Ability; *Writing Skills

ABSTRACT

This paper describes an experiment that was undertaken to examine the usefulness of the cloze test as an objective measure of a native speaker's writing ability. A modified version of the cloze test used by Oller and others to measure integrative language skills in non-native speakers was given to 100 freshman English students. The test incorporated the following features: (1) a procedure of selective deletion; (2) a requirement that the responses fit not only the immediate sentence context but also the overall tone of the passage; (3) the possibility of three credit values for each response: full, partial, or none; and (4) the use of a mathematical weighting procedure to distinguish different slot values. These modifications permitted fine distinctions relating to syntactic fit, semantic fit, and coherence/unity fit, (overall organization). In addition to the cloze test, the students were required to write an essay on a topic related to that of the cloze passage. The scores showed no significant correlations although the ranking order on the cloze tests matched the ranking order for the essays. It was hypothesized that the cloze test was measuring verbal ability, not overall writing ability. Future plans include comparison and analysis of the students' cloze scores with their standard placement test reading and vocabulary scores. Samples of cloze passages and student essays are attached. (Author/AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166974

Measuring writing ability with the cloze test is not closed

by

Helmut Esau

and

Carlson Yost

Department of English
Texas A&M University
College Station, TX 77843

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Helmut Esau
and Carlson Yost

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

FL010094

Measuring writing ability with the cloze test is not closed

0. Introduction. 'Of all the research problems which we face', writes E. D. Hirsch in his The Philosophy of Composition, 'the one most deserving of our attention is the problem of writing assessment' (1977:11). Objective tests are now successfully used to measure reading comprehension and language proficiency in second language situations. Yet no one has found a valid and reliable way of measuring a native speaker's writing ability objectively. We propose to examine here the usefulness of the cloze test as an objective measure of writing ability. First, we introduce a number of modifications into the cloze test as it is currently used. Next we present the results of an experiment in which a cloze test was given to 150 students. Finally, we will assess our empirical results in an attempt to determine what it is that the cloze test measures.

1. Modified cloze test. The cloze test was first introduced by Taylor (1953) to measure the difficulty of reading materials. To construct it, one simply selects a prose passage and deletes every fifth, sixth, or nth word. The testee is then asked to provide a clozure for each blank by inserting the missing word. Since a slight reduction of redundancy within the language code can significantly affect the performance of the non-native speaker, it is not really surprising that the cloze test has been successfully used to test language skills in second language teaching. According to Oller (1973) 'one of the most promising types of integrative skills tests which has been proposed for measuring either achievement or proficiency in foreign language or second language situations is the cloze test' (p. 192). Our own research with approximately thirty different cloze tests during the last three summers has convinced us that it must also be possible to measure the writing skills of native speakers

using the cloze format.

To help increase the validity of the cloze test when used to measure native speakers' writing ability, we have made the following modifications:

(1) Instead of deleting every nth word, we have initiated a procedure of of selective deletion;

(2) we have assigned each response one of three values: full, partial, no credit;

(3) we have developed a mathematical evaluation or weighting procedure to distinguish different slot values, based on the number of correct responses per slot for a given test population;

(4) we have demanded that each response not only fit the immediate or local context but also the overall sense of the passage or global context. With these modifications, the cloze test should make it possible to make some rather fine distinctions relating to semantic fit, syntactic fit, and coherence/unity fit.

2. Experiment. We prepared a cloze test based on selective deletion by removing twenty five words in a prose passage. The items were chosen by trial and error but with the intent of including the entire range from difficult to relatively easy. Below we will illustrate how that distribution can be achieved. The test, which we have reproduced as Appendix I, was then administered to one hundred fifty (150) first and second semester freshman English students at Texas A&M during 1977 and 1978.

We adopted the following evaluation procedure. First, the responses for each slot were tabulated. These were then independently evaluated by three experienced graders of compositions, who were asked to assign each item one of three values: full (=2), partial (=1), or none (=0). The graders met immediately thereafter to discuss the sources of their disagreements. If these could not be

resolved, we simply took them into account in our weighting of the responses. Hence the value for a given response could range from 0 to 6. A response value of 0 meant that all three graders rejected the item, whereas a unanimous full-value choice resulted in a response value of 6.

To calculate the value of each individual slot, we first added up the full value equivalents for the responses to a given slot. Next we divided the number of full-value responses into the number of respondents. This gave us the value of each slot. The procedure enabled us to weight the different slots according to their degree of difficulty. If all responses to a given slot received full credit, the slot value would be exactly one (1) for a full-credit response; if there were only fifty full values out of one hundred fifty responses, the slot value would be three (3); and so on.

The fifth column of Appendix II lists the slot values for all twenty-five slots of the cloze test. The slots have been broken down by parts of speech to show the range of difficulty for the different parts of speech. For instance, the slot value 1½ indicates that slot 4 was the easiest verb clozure, whereas slot 21, with a value of 5, was the most difficult verb clozure in the passage. In general, we were trying to achieve as much spread as possible among the slot values for a particular part of speech. This goal is best realized for verbs and nouns. An inspection of the other parts of speech, particularly conjunctions and prepositions, shows that our cloze test may have had too few items, a problem that we hope to correct the next time we repeat the experiment.

A brief explanation may be needed as to why we did not demand as responses the exact words that had been deleted in the original passage. Such a procedure would have been more objective and hence increased the reliability of the test. However, we do not believe that the results would have taught us very much about the testees' writing ability. In most cases, our judges considered large numbers

of distinct word responses partially or fully acceptable, the maximum being thirty for slot 5. The great variety of word responses for the different slots may be seen in columns 2 and 3 of Appendix II. The second column lists the number of distinct word responses given by our 150 respondents. The third column shows how many distinct words were deemed acceptable by at least one of the three judges.

After we had thus assigned each slot a value, we used those values to score each respondent's test. The fifth column of Appendix II gives the full values for each of the 25 slots. By adding up these values, we determine that the highest possible score that could have been reached is 95. Out of our test population of 150 respondents, the highest score was 52½ and the lowest score 8.

3. Analysis and evaluation. In addition to the cloze test, every student also wrote an essay on a topic related to the cloze passage; some essays were for placement and some were for the student's composition class. The essays were evaluated by at least two independent readers and, in cases of disagreement, by a third reader as well. Each essay was assigned one of four grades: H (high pass), P (pass), M (marginal pass), and F (fail). To determine how well the cloze test measures writing ability, we examined the data for possible correlations between the cloze scores and the essay grades. To our disappointment there were no significant correlations.

We found these results even more puzzling because, with an earlier, 100-item cloze test, there appeared to be a rather high correlation between a student's cloze score and his essay grade. Could it be that our cloze test was too short? That was obviously a possibility. But when we examined the essays individually, we became more and more convinced that we were indeed measuring students' writing ability with the cloze test. To illustrate this fact we have reproduced portions of four student essays in Appendix III. Of these four essays, only D was assigned a passing grade by the two readers, whereas the other three essays (A-C) failed.

However, whenever we asked readers to rank the four writing samples, there seemed to be complete agreement: they always ranked the authors' writing ability from the lowest to the highest as A followed by B followed by C followed by D. When we checked the cloze scores for the four writers, we found the same progression: A scored 9½, B scored 38, C scored 41, and D scored 52½. Moreover, the student who had the highest cloze score wrote also one of the two essays that were ranked highest. Conversely, one of the students who scored lowest on the cloze test wrote also one of the poorest, if not the poorest, essay. The problem is not that we found no patterns but rather that the patterns had too many exceptions.

These observations seemed to suggest that we should perhaps look for correlations elsewhere. We were obviously not measuring the students' overall writing ability, but rather some quantity like verbal ability, which is necessary but not sufficient to write a passing essay. We, therefore, intend to compare our cloze scores with the SAT Verbal (both reading and vocabulary), the SAT Quantitative, and the Nelson-Denny reading and vocabulary scores. Unfortunately, we were unable to complete this part of the experiment in time for this presentation.

Finally, let us make a few general observations that illustrate the diagnostic value of the cloze test. An examination of Column 4 of Appendix II shows that there are relatively few purely syntactic errors. Except for slots 7, 8, and 24, students seemed to have no difficulty identifying the part-of-speech category to which an item belonged. Interestingly enough, two of the three slots that gave students difficulties were transition words, which composition texts generally identify as a problem for beginning writers. The major writing problems, according to our cloze test results, do not occur at the level of sentence syntax but rather at the level of collocational groupings on the one hand and at the point of transition between sentences on the other.

Appendices

(I) Instruction: Read through the following passage. Each numbered blank represents a single word that has been omitted. Your job is to supply a word that will fit. Some of the blanks will have one correct answer, some will have many possible correct answers. Each of your answers should fit the sense and tone of the passage. Write your answers legibly following the numbers provided (suggested maximum time--25 minutes.)

In society's __1__ search for absolute truths and __2__ values, through which the __3__ of living becomes easier to __4__, it has continuously to __5__ a compromise between what is __6__, what seems to be so, __7__, most important, what it __8__ were so. A compromise must __9__ be sought between Truth and that which is __10__ acceptable, convenient, or just __11__, and from this evolves a body of __12__ which Kenneth Galbraith has called the Conventional Wisdom. There is a __13__ of the conventional wisdom to __14__ the intellectual needs of the layman in all areas of American life from __15__ to the family image, and, in all cases, its archenemy is the free-thinking individual.

It has been observed that the first major __16__ to individual freedom __17__ in the replacement of what Sociologists call the Protestant Ethic, which prevailed in this country __18__ the turn of the century, by the __19__ Freudian Ethic. The Protestant and Freudian ethics reveal two opposing concepts of man: the first sees him shaping his own destiny, being governed by standards he sets himself and by his own conscience, therefore responsible for his own acts; the second sees man ruled by unconscious drives and external pressures, hence not really responsible for his acts. His life is shaped not by himself, but by his socio-economic environment; he becomes a failure--perhaps even a criminal--and not he but society is to blame. Under these two warring influences, social __20__ in the United States has been __21__ by escapes into extremes; part of our difficulty, of course, lies in the __22__ that one __23__ has almost entirely replaced the other __24__ of being reconciled and striking a __25__ with it.

(II)

	slot #	# of distinct word responses given	# of different word responses accepted	#of parts-of-speech errors	full slot values
verbs	4	37	22	0	1 1/2
	5	61	30	0	2
	8	35	7	8	3
	14	43	14	1	2
	17	30	9	2	3
	21	86	27	1	5
nouns	3	54	24	0	3
	12	43	23	0	2
	13	64	14	1	3 1/2
	15	53	13	1	8 1/2
	16	50	11	1	8 1/2
	20	59	14	0	5
	22	13	6	0	1
	23	30	13	0	2
	25	36	3	1	3 1/2

	slot #	# of distinct word responses given	# of different word responses accepted	# of parts-of-speech errors	full slot value
adverbs	9	38	18	0	2
	10	40	22	2	3
	24	37	2	35	6 1/2
adjectives	1	42	15	0	1 1/2
	2	52	28	0	4
	19	56	17	3	4 1/2
conjunction	7	19	2	13	1 1/2
preposition	18	15	7	0	2
Predicate Adj/N	6	24	16	1	2
	11	61	23	3	11 1/2

(III)

A. All my life I have tried to be myself. I have tried to make my own decisions and do what I thought was right, as the Protestant ethic believes. However, possessing the human need for acceptance, I have had to drop my set standards and convert to other methods of living and thinking. As long as I fit in with everyone one else, acceptance and friendship was mine. I was "molded" by social and economical pressures until I was no longer myself. However, once accepted I was allowed to change. After changing I realized I had benefited from the experience.

I grew up as an everyday middle class "American boy." It was not until around ten until the pressure from my piers to be like them, do what they did, and don't do anything different or you won't have any friends, began to make a definate change in myself and my lifestyle. At that age, there was no way to escape the pressure to conform if you wanted friends. Of course, being young and succeptable to almost anything, wanting friends and acceptance, I changed, unconcious of the results. I had been molded IN. I was there to stay or put out "to die." This social experience may sound dramatic and harsh, however it cannot be avoided and the possibility of its termination is unlikely and possibly not for the better. The years were passing and I was growing mentally. I was slowly regaining my person. I still went along and did the normal things, according to society, but I felt inside the ideas and practices that were foreign from my own.

B. The behavior of policemen, both as individuals in society and as protectors of society, stems almost entirely from their sense of duty as policemen, and the pressures therein.

Nearly every child, at one time or another, wishes to become a policeman. This is to be expected since they are our "social heroes," as opposed to military heroes. The policemen are supposed to be people that all of society, especially adults since they are supposedly not as naive as children, can look up to. As the children grow up, however, they realize that being in the police force is probably not as full of glory as they had once thought. In fact, as they become adolescents, many come to dislike most violently anything to do with policemen. This is to be expected since many of these young people when younger wanted to become policemen and bring to a halt the things they now delight in. Specifically, these things are usually such stunts as drinking, taking drugs, and reckless driving.

The effect on a policeman from this dissention of a group that he is protecting must be profound. A policeman, assuming, again, that it was his chosen profession many years before actually entering it, has probably given a lot of thought about the state of social union. He has probably come up with many new ideas that he wishes to use and better his country. In training, however, he soon finds out that there is not much room in the police force for individual ideas. He is, for the most part, merely a puppet being trained to be a more experienced puppet. This can have many effects on him. He can accept it and adopt his lifestyle to it. He can go ahead and go through training and find out he really doesn't want the job after all. OR, he can act as though he accepts it, go through the training and become a policeman. This man probably still has his ideas and plans and is sure that they will work if properly applied. Of course, he has to be the one to apply them...

C. For centuries the surgeon has been hailed as the "giver of life." If the health of a patient was endangered, the surgeon following the standards set by his proffession, would endeavor to renew this health. Today society is asking the surgeon to take a second oath, one that clearly conflicts with the first. This oath requires that the doctor, in some instances, should let his patients die.

As a society, we have formed certain laws and guidelines which have have helped us to survive. Among these was the belief that each man controlled his life and could make of that life what he willed. Here then, lies a conflict of ideas. Many members of our society wish their doctors and surgeons to renounce the ancient Hippocratic oath which required them to follow their conscience and expertise in saving the lives of their patients. Instead, society would intervene, choosing when and how its fellow members would die.

Society, like the humans it is composed of, has made many mistakes. I believe that society is also capable of learning from its mistakes. Yet how much better it would be if, in this case, our society learned before the mistakes were made. How could we, as members in this group, trust ourselves with the precious lives of our friends and loved ones. Condemning a man to death, even if he has the slightest chance for recovery, is a decision that can never be retracted. Is our society capable of bearing this burden?

D. There are two standards by which a person's actions are generally judged, the Protestant Ethic and the Freudian Ethic. The former presents the idea that a person is reponsible for his own actions and acts according to his own convictions and conscience. The latter, the Freudian, states that man is merely a victim of circumstances who cannot be held responsible for his own acts. Under this concept, any wrongs committed by a man are blamed on society and fate, never on himself. Either concept may be applied to justify a person's actions, but an application of both to the same situation brings about an unreconciable paradox.

For example, in the case of an alcoholic, judgement by the Protestant Ethic brings on condemnation for several reasons. First of all, the standards set by the alcoholic are lower than they should be, which is the fault of the alcoholic himself. Since he was responsible for setting his standards, he has one strike against him for this alone. Next, he has been overcome by his love for drink and, instead of shaping his own destiny, lets his bottle do it for him. Thirdly, he can hardly be held responsible for his acts when he is drunk while acting. Another point of attack against the alcoholic comes from the basis of the ethic itself. As is implied by its name, the Protestant Ethic originated in the Church. One of the beliefs of the Protestants was that drunkeness was a sin. Many church's view of drinking differ, in fact very few agree on the legality of drinking, but all agree that drunkeness is a sin. Since one of the major characteristics of an alcoholic is his continuous state of inebriation, he is automatically condemned by the institution that founded the ethic by which

he is judged. Lastly, the alcoholic's motive to drink condemns him because many alcoholics apparently drink to still their consciences. Despite the fact that not all alcoholics have this motive, it is attributed to most and therefore tends to condemn all because they fail to follow the dictates of their consciences....